

Are audits wasting resources by measuring the wrong things? A survey of methods used to select audit review criteria

H M Hearnshaw, R M Harker, F M Cheater, R H Baker, G M Grimshaw

Qual Saf Health Care 2003;12:24–28

Objectives: This study measured the extent to which a systematic approach was used to select criteria for audit, and identified problems in using such an approach with potential solutions.

Design: A questionnaire survey using the Audit Criteria Questionnaire (ACQ), created, piloted, and validated for the purpose. Possible ACQ scores ranged from 0 to 1, indicating how systematically the criteria had been selected and how usable they were.

Setting: A stratified random sample of 10 audit leads in each of 83 randomly selected NHS trusts and all practices in each of 11 randomly selected primary care audit group areas in England and Wales.

Participants: Audit leads of ongoing audits in each organisation in which a first data collection had started less than 12 months earlier and a second data collection was not completed.

Main outcome measures: ACQ scores, problems identified in the audit criteria selection process, and solutions found.

Results: The mean ACQ score from all 83 NHS trusts and the 11 primary care audit groups was 0.52 (range 0.0–0.98). There was no difference between mean ACQ scores for criteria used in audits on clinical (0.51) and non-clinical (0.52) topics. The mean ACQ scores from nationally organised audits (0.59, n=33) was higher than for regional (0.51, n=21), local (0.53, n=77), or individual organisation (0.52, n=335) audits. The mean ACQ score for published audit protocols (0.56) was higher than for locally developed audits (0.49). There was no difference in ACQ scores for audits reported by general practices (0.49, n=83) or NHS trusts (0.53, n=383). Problems in criteria selection included difficulties in coordination of staff to undertake the task, lack of evidence, poor access to literature, poor access to high quality data, lack of time, and lack of motivation. Potential solutions include investment in training, protected time, improved access to literature, support staff and availability of published protocols.

Conclusions: Methods of selecting review criteria were often less systematic than is desirable. Published usable audit protocols providing evidence based review criteria with information on their provenance enable appropriate review criteria to be selected, so that changes in practice based on these criteria lead to real improvement in quality rather than merely change. The availability and use of high quality audit protocols would be a valuable contribution to the evolution of clinical governance. The ACQ should be developed into a tool to help in selecting appropriate criteria to increase the effectiveness of audit.

See end of article for authors' affiliations

Correspondence to:
Dr H Hearnshaw, Centre
for Primary Health Care
Studies, University of
Warwick, Coventry
CV4 7AL, UK;
hilary.hearnshaw@
warwick.ac.uk

The issue of appropriately defining and measuring the quality of health care is central to improving clinical practice,^{1,2} and is a fundamental part of clinical governance.³ Healthcare providers and policy makers actively promote quality improvement methods such as clinical audit in the UK or clinical utilisation review in the USA^{3,4} by investing money in them.

One potentially powerful and widely used method of quality improvement is to establish the extent to which clinical practice complies with identified review criteria. The degree of compliance, or lack of it, highlights areas where improvements can be made. This is the basis of clinical audit. Unfortunately, despite large investments in clinical audit, such exercises do not always result in the intended improvements in patient care.^{5–7} It is important to understand the reasons for this lack of success. Rather than considering the whole process of audit as inadequate, one possible explanation lies in the review criteria used.

Review criteria have been defined as “systematically developed statements that can be used to assess the appropriateness of specific health care decisions, services and outcomes”.⁸ Research literature provides guidance on systematic methods of selecting review criteria.^{9–11} As for systematic literature reviews, being systematic is not just searching wider,

but being critical and reducing bias.¹² These systematic methods include the use of high quality research evidence combined with expert consensus judgements to select criteria that are prioritised according to the strength of the evidence and their impact on health outcome. Using systematically selected criteria should increase the likelihood that the quality improvement process will lead to improvements in outcomes of care, rather than just merely changes in structure or process. If performance targets are set according to appropriate criteria, the attainment of these targets should result in improved care. In contrast, if quality of care is assessed against inappropriate criteria, then resources may be wasted on making changes which are unlikely to result in improvement. This would explain some of the lack of effectiveness of clinical audit which has led to lack of commitment to audit activities and clinical governance.

Little is known about the extent to which practitioners actually apply systematic methods to select review criteria for quality improvement. The study reported here is one part of a larger project¹³ and aimed (1) to identify how people conducting quality reviews in the National Health Service (NHS) in England and Wales select their review criteria and (2) to measure the extent to which a valid systematic approach is used.

METHODS

A questionnaire was developed for completion by the lead person in ongoing audits in a random sample of NHS trusts and general practices in England and Wales in 1999. An “ongoing audit” was defined as one in which “the first data collection has begun, or is complete, but the follow up data collection has not finished, and it is less than 12 months since the beginning of the first data collection”. Thus, details should be easily and accurately retrievable for these active audits. Audit leads were contacted via the audit coordinators in NHS trusts or via primary care audit groups.

Questionnaire development

A definition of a systematic method of selecting high quality review criteria was developed from an expert consensus process¹⁴ and used to design a questionnaire to measure how systematically the review criteria had been selected. This was the Audit Criteria Questionnaire (ACQ). The questionnaire asked for the title of the audit and the disciplines it covered, and whether the criteria were drawn from published protocols, published guidelines, selected with the help of audit support staff, or selected by the individual audit lead. A further 25 questions covered whether each element of the systematic approach was considered in their criteria selection (box 1). Items in this list had all been judged by the experts as both important and feasible. Thus, all sets of review criteria should be able to attain a maximum score. Open questions asked about problems experienced in the selection of the review criteria and strategies employed to overcome such problems. The questionnaire was piloted with 37 audit leads from NHS trusts and general practices in England and Wales.

Participant recruitment

A sample size of 270 completed questionnaires was required to show a difference in mean scores of 2% between two categories of responders with 90% power at a 5% significance level. We anticipated a final response rate of about 20% from a multi-level recruitment strategy¹⁵ described below. The random stratified sample comprised invitations to 210 NHS trusts in England and Wales reflecting the overall distribution of trust types (acute, mental health, community, and combinations of those) and 35 primary care audit groups.

Questionnaire distribution

Eighty three audit coordinators from NHS trusts (40% of 210 invited against an anticipated response rate of 50%) agreed to distribute the questionnaire to 10 audit leads, and 11 audit coordinators from primary care audit groups (31% of 35 invited) agreed to distribute the questionnaire to all their practices—that is, not only to those known to have an ongoing audit.

Packs containing a questionnaire, covering letter, and reply paid envelope were provided to audit coordinators. All questionnaires were given a code number. The coordinators were also provided with a copy of the questionnaire for their own reference. In total, 1384 questionnaires were sent out, 830 to NHS trusts and 554 to general practices. Reminders for non-responders were sent 3 and 6 weeks after the distribution of the questionnaire.

Scoring the questionnaire

For each question a score of 1 was given to each “yes” response, 0.5 for “partly”, and 0 for “no”. The score for “partly” was included after the pilot survey showed that a straight dichotomous “yes/no” answer was not always answerable. The response “don’t know” was also scored 0 because, if the information was not known, it could not contribute to the review criteria selection process. The question on how up to date the literature review was scored 1 for <1 year old, 0.5 for 1–4 years old, and 0 for >4 years old. A single rating was constructed by summing the scores for each item and,

Box 1 Questions from the Audit Criteria Questionnaire (ACQ) which were used to calculate the ACQ score

- Were the audit criteria based upon:
 - searching the research literature?
 - consultation with experts?
 - consultation with patients or carers?
 - criteria used in previous audits?
- How up to date was the literature review?
- Was the following information recorded (by you or the authors of the review):
 - the sources/databases used to identify the literature?
 - whether the validity of the research was appraised?
 - the methods used to assess validity?
- Is the method of combining evidence from literature and expert opinion made explicit?
- Is the method used to select the audit criteria described in enough detail to be repeated?
- Were the audit criteria pilot tested for practical feasibility?
- Were the audit criteria prioritised on:
 - impact on health outcome?
 - quality of supporting evidence?
- Were the relative values of harms and benefits associated with treatment options considered in selecting criteria?
- Do the criteria
 - state the patient populations to which they apply?
 - state the clinical settings to which they apply?
 - give clear definitions of the variables to be measured?
 - use unambiguous terms?
- Are the criteria linked to improving health outcomes (rather than, say, to reducing costs or increasing throughput)?
- Do the criteria enable you to differentiate between appropriate and inappropriate care?
- Did the criteria have information on:
 - how the demands of the audit on patients might be minimised?
 - how the demands of the audit on staff might be minimised?
- Did the criteria have clear instructions for using them?
- Were patients consulted about the acceptability of these criteria for them?
- Were all relevant staff consulted about the acceptability of these criteria for them?

because for some items the response option “not applicable” was available, dividing the sum by the number of applicable questions. All questions were thus given equal weighting. This was called the ACQ score with a maximum value of 1 and minimum of 0. It was deemed feasible, as well as desirable, by the expert panel that all sets of review criteria should be able to score the maximum of 1.

Questionnaire characteristics

The validity of the ACQ score was checked using the results from the main study. Internal consistency was confirmed by testing that total scores for the selection of criteria were reliable indices of individual items (Cronbach $\alpha=0.78$). Criterion validity was confirmed by finding that scores for criteria drawn from published sources were higher ($p=0.000$) than those from unpublished sources. All items were answered by at least one respondent. The aspects of review criteria contained in the questionnaire were based upon a list validated by expert opinion and thus had high content validity.

Qualitative data coding

The open questions on problems experienced with selecting review criteria and strategies to overcome problems generated qualitative data. Three members of the research team independently generated and then agreed a coding scheme to categorise the responses from transcripts. All responses were

Table 1 Mean ACQ scores according to audit category

Category	No	Mean ACQ score	95% CI
National	33	0.59	0.53to0.66
Regional	21	0.51	0.41to0.61
With other trust /practice	77	0.53	0.50to0.57
Single organisation	335	0.51	0.49to0.52
All	466	0.52	0.50to0.53

Table 2 Questionnaire items which were absent for more than 40% of respondents (n=82) whose review criteria were derived from published audit protocols

Questionnaire item	% of respondents with this item absent
Did the criteria have information on how the demands of the audit on staff might be minimised?	56
Were patients consulted about the acceptability of these criteria for them?	48
Were the audit criteria based upon consultation with patients or carers?	45
Was it recorded (by you or the authors of the review) whether the validity of the research literature was appraised?	41

then coded according to the developed scheme into a “NUD*IST NVivo” data file to enable content analysis. Three way Cohen’s kappa statistics were calculated to assess inter-rater agreement and ranged from 0.71 to 1.0 across all questions.

RESULTS

All 83 trusts returned at least one completed questionnaire of the 10 distributed (mean 4.60, range 1–10, n=391), giving a response rate of 47%. Completed questionnaires were also received from at least one practice in all of the 11 areas where the primary care audit group had distributed questionnaires (mean 7.73, range 1–21, n=85). The response rate for the general practices is unknown since the number distributed is not known. Ten questionnaires were returned but inadequately completed, giving a total of 466 usable responses.

The mean ACQ score over 466 completed responses was 0.52 (range 0.0–0.98).

Types of audit

From the title, 208 audits (45% of 465 responses giving titles) were classified by the researchers as non-clinical and 257 (55%) as clinical. The classification was done independently by two researchers and then any discrepancies resolved by debate. Between subject univariate analyses of variance showed no difference ($p=0.466$) between ACQ scores for clinical (mean (SD) 0.53 (0.16)) and non-clinical (0.51 (0.17)) audits.

Scope of audit

Audits were reported by respondents as national, regional, run with other trusts/practices in the area, or limited to a single organisation. The mean ACQ scores associated with audits of each type are shown in table 1. Scores from national audits were higher than for any other, but the disparity between frequency of respondents in each category of audit precluded the use of inferential statistical tests of this difference.

Sources of criteria

There was a difference between mean ACQ scores for review criteria drawn from published audit protocols or guidelines (0.56, n=205) and unpublished ones (0.54, n=265), $p=0.003$. The items most often absent from published protocols were investigated and table 2 lists the items absent from more than

Table 3 Mean ACQ scores in general practice and NHS trust settings

	No	Mean ACQ score	95% CI
General practice	83	0.49	0.45to0.52
NHS trusts	383	0.53	0.51to0.54
All	466	0.52	0.50to0.53

Table 4 Frequency of comments from 150 respondents associated with problems in selecting review criteria

Theme	Total
Validity issues	114
Organisational issues	62
Demand issues	47
Literature issues	42
Audit focus	36
Practical issues	35
Attitudes and perceptions	28
Standards issues	23
Total	387

40% of respondents. These indicate where changes in the methods of review criteria selection would be most likely to produce improvements.

NHS trust or general practice setting

There was no significant difference between the mean ACQ scores for review criteria from NHS trusts and general practices (table 3).

Problems and solutions

150 (30%) respondents provided 387 comments identifying problems in selecting review criteria. The eight themes revealed during content analysis are shown in table 4.

Validity issues covered the clarity of the review criteria, whether the criteria were viewed as appropriate, the sample used in the audit, and the quality of data drawn from the

audit. Solutions included recognising the need to make criteria more explicit, ensuring that staff involved in the audit understood which cases should be included, and using an alternative source to verify gathered information.

"Team of auditors took 10 sets of notes each—each auditor tended to put their own spin on the questions."

"Some definitions of the variables to be measured were too imprecise, so we weren't sure that the correct information was collected."

Coordinating different groups of staff in setting up an audit and agreeing review criteria was the most frequently mentioned organisational issue. Reported solutions focused on the value of establishing regular formal meetings.

"Difficult getting staff from two trusts and three departments within trust together, plus getting agreement of guidelines."

Demand issues concerned time and funding limitations, for which more staffing was the only solution offered. Most of the comments around the theme of literature issues concerned a lack of available literature upon which to base criteria, either because of a scarcity of literature on a particular clinical topic or because of the lack of an evidence-based approach for a given clinical discipline. The solutions centred on consultation with colleagues or experts in the area in order to overcome gaps in the evidence base. However, some respondents reported selecting review criteria without such consultation.

"Used my own common sense."

"Made up our own."

Problems with accessing literature included physical access to libraries and problems locating identified publications. There were no solutions reported for solving problems of access.

Being able to refine criteria sets to focus the audit was seen as important in ensuring that the audit was carried out easily. Practical issues included access to the data required for the selected review criteria and lack of adequately skilled/trained staff to collect the data. Solutions suggested were to organise training sessions for staff involved and perseverance.

"Access to a library. Our nearest library is 23 miles away and we don't have any virtual library connections."

The problems with motivation related to audit in general rather than to criteria selection. Solutions suggested were to generate enthusiasm for the audit project by identifying a lead person to maintain communication and enthusiasm. Standards issues included whether the target standard (against the criteria) was realistic.

"Because we developed them locally, setting the standard was probably most difficult—was 100% compliance unrealistic?"

DISCUSSION

We have found that methods of selecting review criteria were often not as systematic as good practice requires. The mean ACQ score for criteria in this study was 0.52 on a scale ranging from 0 to 1. A full score of 1 had been deemed feasible by a consensus of experts so it can be concluded that most respondents had not selected criteria in a systematic, evidence-based way.

A few between group analyses were made. These analyses were useful in determining whether particular types of audit or particular sources of criteria or settings for audit related to the quality of the review criteria. The finding that there was no difference in ACQ scores for clinical and non-clinical audits was perhaps surprising. It might have been expected that audits on non-clinical topics, such as service pathway or organisational structure, would have lower ACQ scores than those examining clinical issues since the evidence base on these topics might be less accessible and less familiar to those conducting audit than for clinical topics. However, this was not the case.

Most of the audits were restricted to a single organisation and were associated with lower scores than review criteria from national or regional audits. However, the rarity of

national or regional audits prevented any inferential analysis of this effect. Nevertheless, the rarity itself is an important result. Given the regional structures for earlier investments in audit programmes in England and Wales, it is disappointing that regional audits were not more widely implemented.

Rather than using those drawn from published audit protocols which had higher ACQ scores, 40% of respondents reported selecting their own criteria. This immediately suggests that a substantial proportion of clinical audits could be improved by the provision and use of published audit criteria. It is encouraging that the programme of the National Institute of Clinical Excellence (NICE) includes the rigorous selection of both clinical guidelines and audit protocols.¹⁶ However, ACQ scores for criteria based on published protocols still only achieved a mean score of 0.59 so, even if the published protocols were based on good evidence, the protocol usually did not provide enough accompanying information about its development. This is an important issue because, if published audit protocols do not provide complete details of how their review criteria were selected, it is almost impossible to make an informed choice on their appropriateness. In order to ensure that criteria are valid, it is necessary to know their evidence base, the quality of that evidence, and the reasons behind any prioritisation. This, for example, is now a standard expectation of clinical guidelines.¹⁷ We should therefore expect published audit protocols to include a detailed and transparent account of how they were selected. Our results show that this is not the case for many of the published audit protocols used in England and Wales. In particular, the omitted items such as involvement of patients in selection of review criteria, information on the demands of criteria on patients and staff, and the report of any evaluation of the validity of the literature would be relatively easy to provide.

A number of problems were identified to explain why scores for the selection of the review criteria were generally low. Several respondents commented on difficulties in locating or gaining access to the literature: some respondents had trouble in narrowing down large criteria sets to produce a manageable audit protocol. Criteria backed by up to date valid research evidence and piloted would reduce such problems. Effective access to literature, which should be relatively easy for NHS organisations to provide, would enable these processes.

Organisational problems such as the coordination of meetings to discuss criteria selection or the amount of time taken to set up the audit and select criteria could mean that short cuts are sometimes taken in the criteria selection process. For some respondents an evidence-based approach had not yet taken hold in their discipline, with little research evidence to guide their practice. In some cases, individuals preferred to rely on their own "common sense" to select criteria in the absence of published evidence. This could be considered a risky strategy.

Many respondents were aware that they could have been more systematic in their approach to selecting review criteria, but they had faced obstacles. In many cases they were also aware of strategies to overcome these obstacles. Thus, although our results show that there are serious problems impeding the selection of appropriate review criteria for audit, they also show that solutions can be identified.

The sample of respondents in this study succeeded in producing a study power large enough to justify generalising our results to most clinical audits in the NHS. Any bias in our sample of volunteer responders would be towards those who felt their performance was good enough to report to the researchers. We can therefore infer that ACQ scores for non-responders would be even lower than those of the responders. This reinforces our concern with the low quality of review criteria selection.

The ACQ provided a valid tool to assess the quality of methods used to select review criteria for clinical audit. The creation of this instrument has important implications for those

Key messages

- If quality of care is assessed against inappropriate review criteria, resources may be wasted in ineffective quality improvement activities.
- It is possible to measure the quality of review criteria selected for audits.
- Many audits have used review criteria which are not well selected.
- Problems in criteria selection include difficulties in coordination of staff to undertake the task, lack of evidence, poor access to literature, poor access to high quality data, lack of time, and lack of motivation.
- Potential solutions include investment in training, protected time, improved access to literature, support staff, and availability of published protocols.
- The availability and use of high quality audit protocols would be a valuable contribution to the evolution of clinical governance.

evaluating quality improvement programmes. Identification of good practice in criteria selection should enable strengths to be built on and identification of less ideal practice may suggest remedial measures for future quality improvement activities. Furthermore, the scoring method of the ACQ allows identification of the aspects of the criteria which are missing. This may help audit practitioners when selecting between previously published sets of audit criteria.

Audits could be much more effective and investment in audit as part of clinical governance more cost effective if criteria were selected more carefully. We have identified ways to enable better criteria selection. This would make audits more likely to lead to improvements without increasing time and effort. Published usable audit protocols providing evidence-based review criteria with information on their provenance will provide a valuable contribution to the evolution of clinical governance. Although this study was conducted in England and Wales, these conclusions can be applied to other countries where quality review is practised since the criteria developed here for the assessment of review criteria are relevant to all reviews.

ACKNOWLEDGEMENTS

The authors are grateful to the audit coordinators who distributed the questionnaires and the audit leads who completed and returned them. We are also grateful to the anonymous reviewers for constructive and encouraging comments.

Authors' affiliations

H M Hearnshaw, Centre for Primary Health Care Studies, University of Warwick, UK
R M Harker, National Children's Bureau, London, UK

F M Cheater, School of Healthcare Studies, University of Leeds, UK
R H Baker, Clinical Governance Research and Development Unit, Department of General Practice and Primary Health Care, University of Leicester, UK
G M Grimshaw, Centre for Health Services Studies, University of Warwick, UK

The study was supported by the UK NHS R&D Health Technology Assessment programme. The views and opinions expressed here are those of the authors and do not necessarily reflect those of the NHS Executive.

REFERENCES

- 1 **Walshe K**, Spurgeon P. *Clinical audit assessment framework handbook*. Series 24. Birmingham, UK: Health Services Management Centre University of Birmingham, 1997.
- 2 **Harvey G**, Kitson A. Achieving improvement through quality: an evaluation of key factors in the implementation process. *J Advan Nursing* 1996;**24**:185-95.
- 3 **Department of Health**. *A first class service: quality in the new NHS*. London: HMSO, 1998.
- 4 **Department of Health and Human Services**. *The challenge and potential for assuring quality health care for the 21st century*. Publication No OM 98-0009. Prepared for the Domestic Policy Council, Washington DC web page, 17 June 1998 (available at <http://www.ahcpr.gov/qual/21stcena.htm>).
- 5 **Berger A**. Action on clinical audit: progress report. *BMJ* 1998;**316**:1893-4.
- 6 **Scally G**, Donaldson LJ. Clinical governance and the drive for quality improvement in the new NHS in England. *BMJ* 1998;**317**:61-5.
- 7 **Davis DA**, Taylor-Vaisey A. Translating guidelines into practice: a systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines. *Can Med Assoc J* 1997;**157**:408-16.
- 8 **Field M**, Lohr KN. *Guidelines for clinical practice: from development to use*. Washington: Institute of Medicine, 1992.
- 9 **Campbell SM**, Roland MO, Shekelle PG, et al. Development of review criteria for assessing the quality of management of stable angina, adult asthma, and non-insulin dependent diabetes mellitus in general practice. *Qual Health Care* 1999;**8**:6-15.
- 10 **Fraser RC**, Khunti K, Baker R, et al. Effective audit in general practice: a method for systematically developing audit protocols containing evidence-based review criteria. *Br J Gen Pract* 1997;**47**:743-6.
- 11 **Ashton CM**, Kuykendall DH, Johnson ML, et al. A method of developing and weighting explicit process of care criteria for quality assessment. *Med Care* 1994;**32**:755-70.
- 12 **Petticrew M**. Systematic reviews from astronomy to zoology: myths and misconceptions. *BMJ* 2001;**322**:98-101.
- 13 **Hearnshaw HM**, Harker R, Cheater F, et al. A study of the methods used to select review criteria for clinical audit. *Health Technol Assess* 2002;**6**(1).
- 14 **Hearnshaw HM**, Harker R, Cheater F, et al. Expert consensus on the desirable characteristics of review criteria for health care quality improvement. *Qual Health Care* 2001;**10**:173-8.
- 15 **McAvoy B**, Kaner E. General practice postal surveys: a questionnaire too far? *BMJ* 1996;**313**:732-3.
- 16 **National Institute of Clinical Excellence (NICE)**. <http://www.nice.org.uk> (accessed January 2001).
- 17 **Graham ID**, Calder LA, Hebert PC, et al. A comparison of clinical practice guideline appraisal instruments. *Int J Technol Assess Health Care* 2000;**16**:1024-38.